Structure *14*, 1341–1344, September 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.08.003

A DNA-Centric Look at Protein-DNA Complexes

The availability of many DNA-protein structures makes their classification timely and important. In this issue of *Structure*, the method of Akinori Sarai and his collaborators (Prabakaran et al., 2006) utilizes aspects of the binding interactions and DNA properties to identify seven clusters of structures with a classification scheme that differs significantly from previous approaches.

The recent growth in the number of experimental protein structure determinations, together with advances in computational modeling, is rapidly providing an increasing number of structures of protein-protein and protein-nucleic acid complexes. Even the structures of extremely large assemblages of proteins and nucleic acids, such as the ribosome, have been successfully determined experimentally (reviewed in Ramakrishnan, 2002).

The situation in the field of protein-DNA complexes now resembles the situation with protein structures a decade ago. The large body of known protein structures then led to the development of extremely useful protein structure classifications such as those provided by SCOP (Murzin et al., 1995) and CATH (Orengo et al., 1997). The accumulation of many solved structures of protein-DNA complexes creates the possibility of doing the same for these complexes.

The traditional way in which protein-DNA structures have been classified is based on the structural characteristics of the proteins, and particularly on those parts which interact with the DNA (Luscombe et al., 2000). The story began long ago with the famous helix-turn-helix structure (Anderson et al., 1981), initially considered to be "the" motif for protein-DNA interactions, followed over many years with the accumulation of other interaction motifs, usually based on the protein structure alone. Now, Akinori Sarai and his collaborators take a new direction from a new point of view (Prabakaran et al., 2006). In their study, they move away from a "protein-centric" viewpoint, and provide an entirely new classification scheme based on 11 structural descriptors relying principally on interaction considerations and the bound DNA characteristics. Initially, they considered a larger set of 22 characteristics, but pared this list to the ones providing the most meaningful characterization. The interaction factors include the number of contacts and the buried surface area, the protein contact order, as well as z-score values based on empirical DNA-protein potentials, representing the specificities of protein-DNA recognition. The DNA properties include the major and minor groove depths and widths, the DNA helical bend angle, the similarity of the DNA to both A- and B-form DNAs, and the GC-content of the DNA. Their cluster analysis is thorough, using a number of statistical tests and two different types of clustering, and the resulting classification is compared with those worked out in the past by other researchers (for example, Luscombe et al., 2000). The final result of their analysis on 62 DNA-protein complexes is the delineation of seven structurally distinct clusters. One of the highly interesting aspects of this work is that proteins previously grouped together by traditional methods of protein structural classification are separated into many different clusters. For example, it is quite striking that helix-turn-helix proteins are found in six of the seven classes. This finding suggests that the three-dimensional details of the protein-DNA interactions and of the changes in the DNA structure in the bound state are truly significant descriptors of the complexes. This supports the viewpoint that distortions in DNA structure as induced by its interactions with the protein are important aspects of the overall binding process and reflect the protein structure itself.

The largest cluster includes 13 structures where the DNA has a characteristically compressed major groove, a somewhat widened minor groove, and a smaller than average interfacial surface area. The proteins themselves in this class are extremely diverse: (1) lac repressor, (2) engrailed homeodomain, (3) MyoD, (4) paired domain, (5) Tc3 transposase, (6) trp repressor, (7) erythroid transcription factor Gata-1, (8) GAGA factor met repressor, (9) Snad MH1, (10) bovine papilloma virus E2 protein, (11) regulatory factor RFX1, (12) Epstein-Barr virus nuclear antigen 1, and (13) regulatory factor RFX1. This structural analysis provides a remarkable link between proteins that would previously have been considered to be completely unrelated to one another. The work is particularly notable in its reliance on DNA characteristics, and it is even noteworthy that such a classification is possible. The likely reason that it is possible is because, even though the DNA double helix is relatively flexible, binding to a specific protein selects among the available conformations, and the selected conformation is reflective of the protein itself. This is perhaps a subtle point, but the approach is clearly successful.

The underlying message is that it can be more important to investigate how the DNA is changed as a result of complex formation instead of focusing on the relatively unchanged protein. This insight and the DNA-centric point of view have led to the present classification scheme, providing a fresh perspective from the older protein-centric classification schemes. Not just a second look, but a second viewpoint can often be useful!

Of course many other properties of protein-DNA structures could be used for classification. An appealing one is to use the dynamics of the complexes. The relationship between the structure and dynamics of biological macromolecules is now well established and can readily be investigated using either elastic network models (Atilgan et al., 2001;Bahar et al., 1997) or molecular dynamics (Villa et al., 2005).

The newly described DNA-centric classification scheme will undoubtedly be thought-provoking for

researchers and will hopefully suggest new experiments to test and expand the approach.

Taner Z. Sen,[1,2] Andrzej Kloczkowski,[2] and Robert L. Jernigan[1,2]
[1] L.H. Baker Center for Bioinformatics and Biological Statistics
[2] Department of Biochemistry, Biophysics, and Molecular Biology
Iowa State University
Ames, Iowa 50011

Selected Reading

Anderson, W.F., Ohlendorf, D.H., Takeda, Y., and Matthews, B.W. (1981). Nature *290*, 754–758.

Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., and Bahar, I. (2001). Biophys. J. *80*, 505–515.

Bahar, I., Atilgan, A.R., and Erman, B. (1997). Fold. Des. *2*, 173–181.

Luscombe, N.M., Austin, S.E., Helen, M., and Thornton, J.M. (2000). Genome Biol. *1*, 1.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). J. Mol. Biol. *247*, 536–540.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). Structure *5*, 1093–1108.

Prabakaran, P., Siebers, J.G., Ahmad, S., Gromiha, M.M., Singarayan, M.G., and Sarai, A. (2006). Structure *14*, this issue, 1355–1367.

Ramakrishnan, V. (2002). Cell *108*, 557–572.

Villa, E., Balaeff, A., and Schulten, K. (2005). Proc. Natl. Acad. Sci. USA *102*, 6783–6788.

Structure *14*, September 2006 DOI 10.1016/j.str.2006.08.002

A New GNAT in Bacterial Signaling?

Acyl-amino acids were discovered in the search for novel therapeutic agents produced by uncultured microorganisms. In this issue of *Structure*, Van Wagoner and Clardy (2006) find that the *N*-acyl-amino acid synthase FeeM bound to acyl-tyrosine is the latest member of the GCN5-related *N*-acyltransferase superfamily.

Acyl-amino acids were discovered in bacteria as biosynthetic products of cloned bacterial operons that had been ectopically expressed in *E. coli* (Brady et al., 2002). However, their functions in vivo remain unclear. A new study by Van Wagoner and Clardy (2006) reveals the structure and mechanism of the enzyme, FeeM, which produces acyl-tyrosine during the process of amino acid enol ester synthesis via the *f*atty acid *e*nol *e*ster (*fee*) operon (Brady et al., 2002, 2004). The authors discuss the role of acyl-tyrosine as an antibiotic and suggest a possible novel role in cell-cell signaling.

The structure of FeeM in complex with a fortuitously bound acyl-tyrosine (Van Wagoner and Clardy, 2006) (Figure 1A) strongly resembles the GCN5-related N-acyltransferase (GNAT) superfamily. This superfamily includes the N-acetyltransferases (Figure 1B) (reviewed in Clements and Marmorstein, 2003; Vetting et al., 2005) and the acyl-homoserine lactone (AHL) synthases (Figure 1C) (Gould et al., 2004; Vetting et al., 2005; Watson et al., 2002). Although there is no significant sequence similarity among these enzymes, the striking structural similarity suggests that FeeM should be considered the newest member of the GNAT superfamily (Vetting et al., 2005) (the rmsd of FeeM compared to GCN5, serotonin-N-acetyltransferase (AANAT), LasI, and EsaI is near 1.65 Å for over 50% of the Cα carbons).

The substrates of the GNAT superfamily members have common characteristics. First, one substrate provides a primary amine, which is acylated by the enzyme. These substrates include lysines in histone tails, serotonin, aminoglycosides, *S*-adenosyl-L-methionine (SAM), and now free amino acids. Second, the other substrate contains a phosphopantetheine moiety in the form of either acyl-acyl carrier protein (acyl-ACP) or acetyl-coenzymeA (acetyl-CoA). Therefore, GNATs appear to have evolved for the purpose of binding to the phosphopantetheine moiety of the substrate (Clements and Marmorstein, 2003).

Interestingly, the enzymatic mechanism of acylation is not conserved among the GNATs (Clements and Marmorstein, 2003). For example, studies of the GCN5 and AANAT enzymes support a mechanism that involves a direct nucleophilic attack of the deprotonated amine on the C1 position of the acyl group (Figure 1E) (Hickman et al., 1999). The amine is deprotonated by a water molecule that is part of a "proton wire" of highly ordered and positionally conserved water molecules in the V-shaped active site cleft between β strands 4 and 5 (Figure 1B). In contrast, the mechanism of histone acetylation by Esa1 proceeds through a covalent thioacyl-enzyme intermediate (ping-pong mechanism) (Figure 1E) (Yan et al., 2002), while the precise mechanism of acylation used by AHL synthases is not known (Figure 1F). FeeM performs an acylation reaction (Figure 1D), which Van Wagoner and Clardy (2006) suggest occurs by the direct nucleophilic attack mechanism.

The proposed mechanism of FeeM is consistent with the putative functions of other genes in the *fee* operon. The operon houses analogs of all of the genes that would be required for the production of the acyl-tyrosine product, including *feeL* itself, which encodes the ACP homolog. Therefore, unlike the other GNATs that scavenge ubiquitous acetyl-CoA or acyl-ACP, FeeM appears to have coevolved with a specialized substrate, acyl-FeeL. In support of this, the authors demonstrate that